THE JOURNAL OF PHYSICAL CHEMISTRY B

Caltech Library

Subscriber access provided by Caltech Library

Article

Geometrical Description of Protein Structural Motifs

John J Kozak, Harry B. Gray, and Pernilla Wittung-Stafshede

J. Phys. Chem. B, **Just Accepted Manuscript** • DOI: 10.1021/acs.jpcb.8b07130 • Publication Date (Web): 24 Aug 2018

Geometrical Description of Protein Structural Motifs

John J. Kozak*[a], Harry B. Gray[b] and Pernilla Wittung-Stafshede*[c]

a. Department of Chemistry, DePaul University, Chicago, IL 60604-6116, United States

b. Beckman Institute, California Institute of Technology, Pasadena, CA 91125, United States

c. Department of Biology and Biological Engineering, Chalmers University of Technology, 41296 Gothenburg, Sweden

* Corresponding authors: kovalik0301@gmail.com; pernilla.wittung@chalmers.se

Abstract

We present a geometrical method that can identify secondary structural motifs in proteins via angular correlations. The method uses crystal structure coordinates to calculate angular and radial signatures of each residue relative to an external reference point, as the number of nearest-neighbor residues increases. We apply our approach to the blue copper protein amicyanin using the copper cofactor as the external reference point. We define a signature termed Δβ which describes the change in angular correlation as the span of nearest neighbor residues increases. We find that three turn regions of amicyanin harbor residues with Δβ near zero, while residues in other secondary structures have Δβ greater than zero: for β-strands, Δβ changes gradually residue-by-residue along the strand. Extension of our analysis to other blue copper proteins demonstrated that the noted structural trends are general. Importantly, a purely geometrical description of the folded protein accounts for all forces holding the structure together. Through this analysis, we identified some of the turns in amicyanin as symmetrical anchor points.

Introduction

Proteins are polypeptide chains that must fold into unique three-dimensional compact structures in order to function. In seminal work published in 1959, Walter Kauzmann first laid out a convincing case that a "hydrophobic effect" operates as the driving force for folding of polypeptides into compact structures in aqueous solution[1,2]. Following this groundbreaking work, influential contributions from Tanford appeared[3-5], as well as a "special Kauzmann issue" on experimental, computational, and theoretical aspects of the hydrophobic effect[6]. More recent contributions in this area include papers by Rose et al.[7], Compiani and Capriotti[8], and Chandler[9].

Many thousands of protein structures have been determined to atomic resolution, motivating theoretical and experimental work that has shed light on the factors that control polypeptide folding processes[10-12]. Remarkably, nowadays it is even possible to design new proteins with specified activities[13], although the design principles are not fully developed. As a multitude of forces governs folding reactions, there is a need to find new ways to determine the effects of these forces on both protein function and evolution. Here we present a quantitative approach to describe protein folded structures via calculation of geometrical signatures. The angular phase diagrams that result from our analysis are similar in spirit to the Ramachandran phi/psi plots[14,15]. In Ramachandran's approach, angular correlations between a given residue and its first nearest neighbors are studied and, by calculating the backbone dihedral angles (φ,ψ) for all residues, regions with no steric conflicts are identified. In our method, angular correlations between extended ranges of neighboring residues are investigated with respect to a reference point *external* to the protein backbone.

Our method is based on earlier work where we explored the stability of the folded state of the blue copper protein amicyanin[16-18]. We then developed a simple geometrical model of an

unfolded state that featured a fully-extended, linear sequence of triplet residues, with the central residue in each triplet having the same angular relation to the copper (Cu) as in the folded state. We calculated the radial displacement of each residue relative to Cu as the protein unfolded (expanded) from the folded structure. Implementation of our geometrical model allowed quantification of the resiliency (resistance to perturbation) of different amicyanin sequence segments, which in turn was related to existing experimental data on amicyanin unfolding. We now focus on *folded* amicyanin, aiming to define in strictly geometrical terms the structure of the folded state. Most importantly, our approach specifically identifies some of the turns in the protein as symmetrical anchor points.

Angular Description of the Folded State

Using classic theorems from trigonometry, we extract information from the structural coordinates in 1AAC for *Paracoccus denitrificans* amicyanin[19]. As noted above, we assign Cu to be the origin of our reference coordinate system. With respect to this "spectator atom", we then calculate the distance from Cu to the α-carbon of each residue i (of the 105 in the protein). Displayed in **Fig. 1A** is the radial α-carbon to Cu distance from each residue with the color code indicating helical, β-strand and unstructured regions as defined in 1AAC.

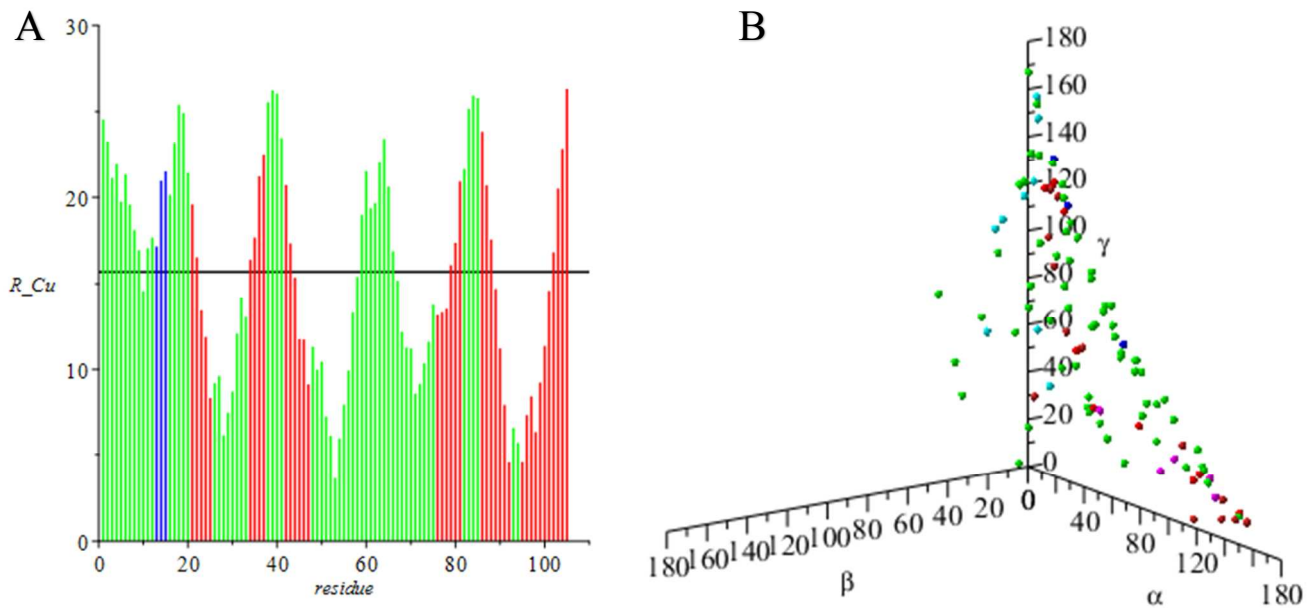

Figure 1. A. Distance (in Å) of residues to Cu; α-helical segments are in blue, β-sheet segments in red, and unstructured regions in green. The horizontal line gives the overall, average displacement of the 105 residues in the protein. **B.** Angle phase diagram for amicyanin for the triplet module. Color convention as in A.

To explain the calculation of angular coordinates, we consider residue i=3 and its nearest neighbors, i.e., a triplet of residues (n=3). The distance from Cu to the α-carbon of residue i=3 is 21.12 Å. The through space distance between the α-carbons of the two residues (i–1 and i+1) flanking residue i=3 in the triplet centered on i=3 is 7.03 Å. Given these distances, we calculate the three angles [β_i, α_i, γ_i], as follows: for the triplet centered on residue i=3, β_i is the angle between lines drawn from Cu to residue 2 and to residue 4; α_i is the angle between the line from Cu to residue 2 and the line connecting residues 2 and 4. Finally, γ_i is the angle between the line

from Cu to residue 4 and the line connecting residues 2 and 4. The angles calculated (via the arccos) are found to be: $\beta(i=3) = 17.61^{o}$; $\alpha(i=3) = 91.71^{o}$; $\gamma(i=3) = 70.67^{o}$. Together, the angles $[\beta_i, \alpha_i, \gamma_i]$ form a triangle (sum to 180^{o}). As another example, consider the extended sequence of five residues (n=5: including i–2, i–1, i, i+1, i+2) centered on residue i=3. Here, the through space distance between α-carbons of the terminal residues (i–2=residue 1 and i+2= residue 5) in the five residue segment centered on residue i=3 is 10.53 Å. The angles $[\beta_i, \alpha_i, \gamma_i]$ in the triangle formed with the Cu at one vertex are: $\beta(i=3) = 24.59^{o}$; $\alpha(i=3) = 104.09^{o}$; $\gamma(i=3) = 51.32^{o}$.

Again, as must be the case, the three angles sum up to 180^{o}. To visualize the concept, triangles with the Cu ion for n=3 and n=5 (indicating the angle β in both triangles) for an arbitrary selected residue, i, is shown in the amicyanin structure in **Fig. S1**.

In the above triplet (n=3) calculation, repulsive and attractive interactions between each residue and its first-nearest neighbors are retained in the geometrical parameters. For comparison, in an *ab initio* approach, the Hamiltonian for atoms making up a particular triplet of residues would be formulated by taking into account all (pairwise) attractive and repulsive interactions between the involved atoms. Employing such a computational quantum chemistry approach, the geometry of the triplet can be (derived and) compared with experimental data. In contrast, in our empirical approach, structural information is directly used to define the geometry of each triplet which thus contains *all* interatomic interactions. That the geometrical representation includes both repulsive and attractive interactions is an important distinction between our approach and the traditional Ramachandran analysis, since the latter[14,15] is based on purely repulsive interactions. In further contrast to the Ramachandran analysis, our method extends beyond first-nearest neighbors (n up to 15 analyzed here) and thus takes into account a multitude of intra-protein interactions.

For the n=3 modular unit, distances and angles are calculated for a triad of residues centered on residue i, and can be applied to residues i=2-104. As the length of the modular unit increases, the number of residues at the C-terminal and N-terminal ends *not* included also increases. For a modular unit of five residues (i–2, i –1, i, i+1, i+2), distances and angles can be calculated for all residues except the two N-terminal residues and the two C-terminal residues. For a segment of 15 residues, the first N-terminal residue for which the above calculations can be carried out is residue i=7. We performed angular calculations for all residues in amicyanin including triplet up to 15-residue segments, subject to the caveat noted above. See **Fig. S2** in Supplemental Information (SI) for plots of β values for all residues as a function of increasing number of nearest neighbors (from n=3 to 15).

Displayed in **Fig. 1B** is the angular "phase diagram" [β_i, α_i, γ_i] for a triplet unit of n=3 with points in blue corresponding to the α-helix, points in red representing the six β-strands, and points in green designating unstructured regions. The representation in **Fig. 1B** is instructive, but not useful. We therefore prepared a projection of the three dimensional representation in **Fig. 1B** onto a two dimensional (α, γ) plane. The projections (β, α), (β, γ), and (α, γ) also have been studied (not shown). However, all the necessary conclusions can be illustrated using the (α, γ) projection (**Fig. 2A**). In the panels of **Fig. 2**, we have connected data points for residues in a given structural region. Blue segments denote the α-helix, red segments denote β-strands, and green segments denote unstructured regions of the protein. The black triangle in **Fig. 2A** is the smallest triangle that defines the boundary within which all (α, γ) values are found for the n=3 (triplet) module analysis of the folded structure. In our approach, this boundary plays the same role as the accessible regions in a Ramachandran φ/ψ plot. As noted previously, angular

correlations in φ/ψ plots are calculated with respect to nearest-neighbor backbone residues only, whereas we here calculate angular correlations with respect to a point *external* to the backbone.

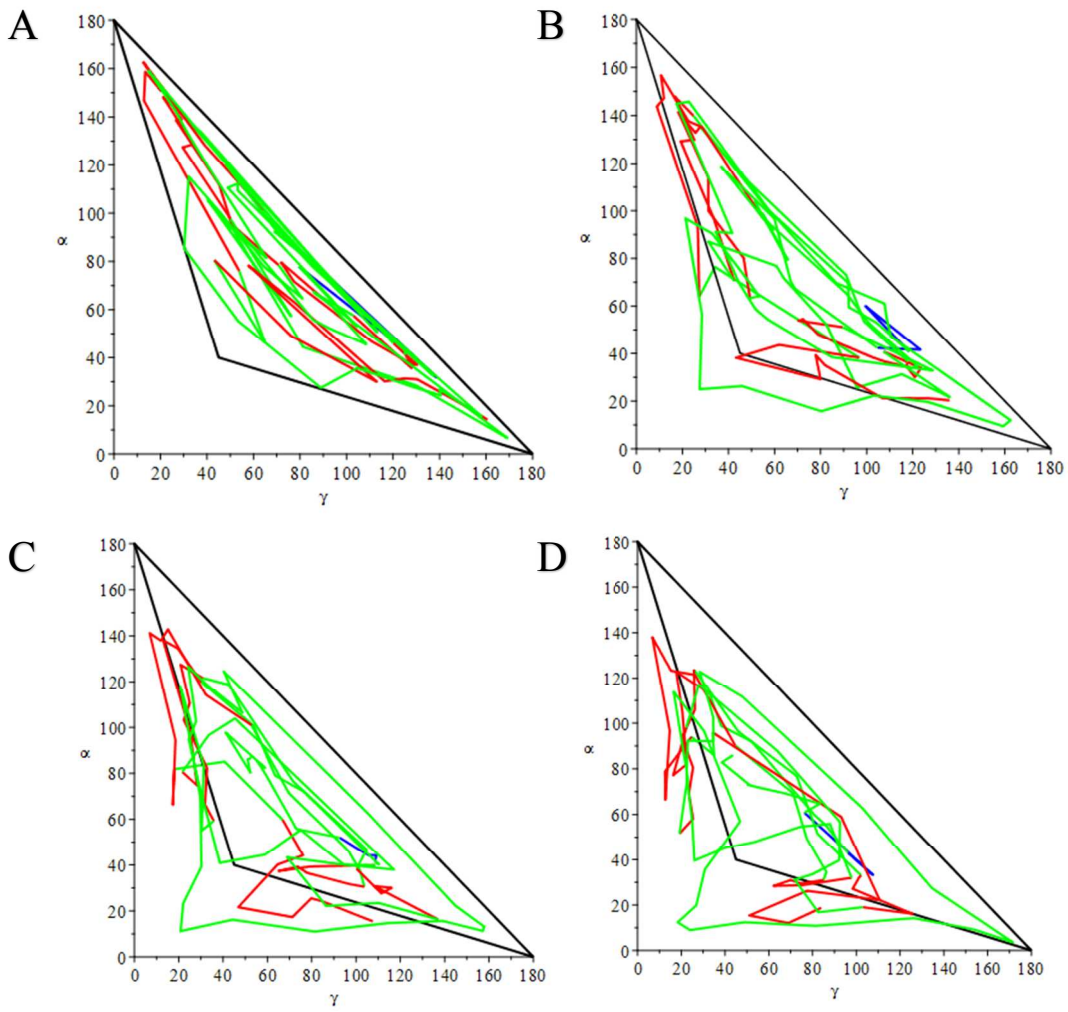

Fig. 2. A. Projection of the angle phase diagram for the triplet module in Fig 1B to the (α, γ) plane. **B.** Projection of the angle phase diagram to the (α, γ) plane for five residue modules. **C.** Projection of the angle phase diagram to the (α, γ) plane for seven residue modules. **D.** Projection of the angle phase diagram to the (α, γ) plane for nine residue modules. Color convention as in Fig. 1.

Defining an Angular Difference Function

Presented in **Fig. 2B** is the (α, γ) projection for five residue segments centered on residue i for all residues in the protein. **Fig. 2B** includes the triangular boundary identified for the triplet modular unit. For five residue segments, most points remain confined within the (triplet) triangular boundary, but one segment, corresponding to residues 47-75, has residues lying outside the boundary. Segments in **Fig. 2B** outside the triplet boundary thus represent configurations accessible to the protein when longer-range correlation effects (here, five-residue segments) are taken into account. This expansion of angular space is further augmented for modular units of seven and nine residue segments and includes both unstructured and β-strand regions (**Fig. 2CD**). Notably, segments external to the triplet boundary triangle all have smaller α and γ values than those displayed in the triangle for triplet modules. Since the three angles must sum to 180^{o}, the β values must be larger for residues in these segments.

Overall trends shown in **Fig. 2BCD** panels are helpful, but we would like to have a more quantitative metric. In the one we have adopted, β angles are used to construct an angular difference function called Δβ (delta beta) that reports on, residue by residue, the angular difference between an n-residue modular unit and the triplet modular unit, each centered on a given residue i. For example, for a five residue modular unit centered on residue 3, Δβ is the value derived for the 5 residue module (i.e., from residue 1 to 5) minus the β value derived for the 3 residue module (i.e., from residue 2 to 4). A plot of Δβ versus residue number for seven-residue modules is shown in **Fig. 3A**. Notably, almost all Δβ values are greater than zero, providing confirmation that the β angle is larger for higher n values. By examining Δβ values for nine-residue segments (**Fig. 3B),** we see that trends are similar between 7 and 9 residue modules. In **Fig. S3**, we show all Δβ plots for n=5 to 15.

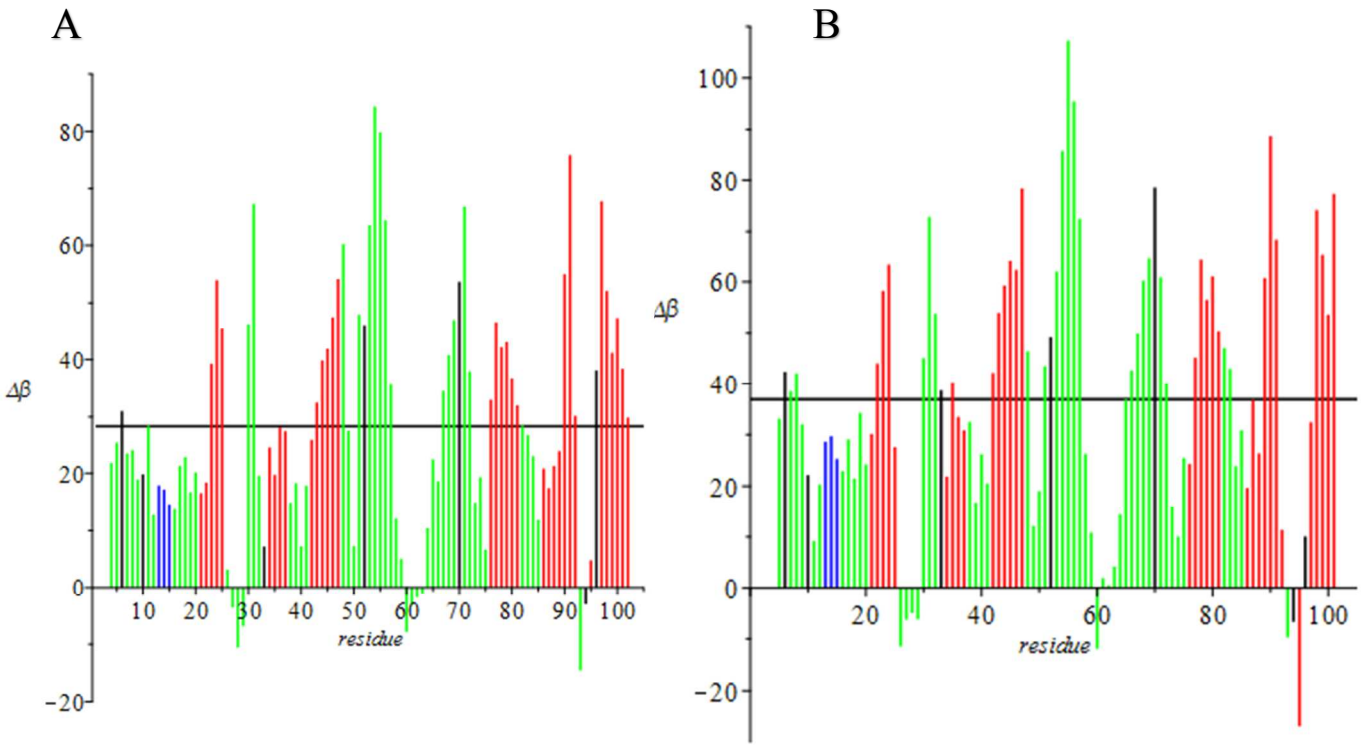

Fig. 3. Angular difference (Δβ) versus residue number for seven (**A**) and nine (**B**) residue modules. The horizontal line in each panel denotes the overall average value of Δβ. Color code as in Fig 1; black denotes proline residues.

Geometrical Signatures of Secondary Structures

Examination of **Fig. 3** shows that angular correlations vary in distinct patterns that reflect geometrical properties around each amino acid. A region of low Δβ is followed by a stretch of high Δβ and so on, in an apparent 'wave' pattern. In a broad sense, these waves reflect the polypeptide backbone winding back and forth toward the external reference point. Notably, several residues around position 60 in the largest unstructured segment (including residues 48 to

75) have Δβ values close to zero in both 7 and 9 segment plots (and for higher n values as well, **Fig. S3**). The same pattern is found around residues 26-29 and 93-94, which also reside in unstructured (i.e., green) regions of the protein.

Residues in a zero Δβ region display an angle relative to Cu that does not change when there is an increase in the number of those on each side of the residue of interest. For β to remain constant, the residues on both sides must lie within the same triangle as the one formed by Cu and nearest neighbor residues. It follows that, with respect to Cu, such a peptide stretch is radially symmetrical. Inspection of the amicyanin structure reveals that the three regions with almost zero Δβ correspond to unstructured areas specifically defined as turns in 1AAC (**Fig. 4A,** purple). In fact, all six identified turn regions in amicyanin have low Δβs, but absolute values vary somewhat, indicating that the exact symmetry with respect to Cu differs among the turns (**Fig. 4A,** purple and green). From a symmetry perspective, turn regions in general should have low Δβ values as the polypeptide 'folds back' on itself at these places. Further, turns with Δβ of around zero serve as distinct symmetrical anchors in the folded structure.

β-strand residues often have high Δβ values, which is expected for extended structures; expanding neighboring residues outward (increasing n) will increase the angle of the residue segment with the Cu center. Concomitantly, for fixed n, when moving along a β-strand residue by residue in amicyanin, the angle relative to the external point gradually increases or decreases (**Fig. 3**). For n=9 and lower values, there is a gradual increase in Δβ values for increasing residue number in β-strands going towards the Cu (strands 1, 3 and 5), whereas Δβ values for β-strands going away from the Cu (strands 2, 4 and 6) gradually decrease as the residue number increases, thereby implying that such values (for a fixed n) are larger closer to the external point. This

directional railing related to the Cu center disappears at n=11 and higher n values.

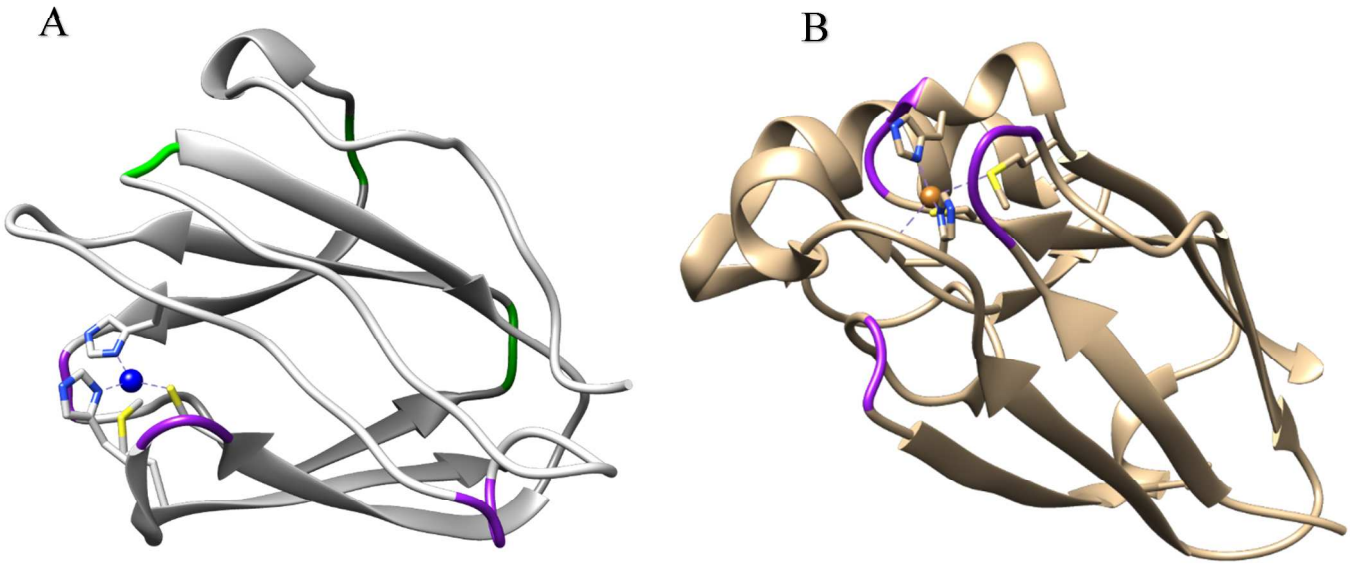

Fig. 4. A. Amicyanin structure (1AAC, grey) with six defined turn regions highlighted in color. In purple are three turns with zero or negative Δβ (residues 27-28, 59-61 and 93-94), see Fig. 3. In green are the other three defined turns that all have low Δβ (residues 18-19, 39-40, and 74-75). **B.** Azurin structure (1JZF, gold) with the three defined turn regions in purple that have zero or negative Δβ (residues 10-13, 89-90 and 114-115), see Fig. S4. The copper is shown in blue in both figures together with side chains of coordinating residues. The two structures are overlaid in Fig. S5 showing that two purple loops overlap structurally.

Method Validation

For comparison, we performed the same analysis on other blue copper proteins, namely rusticyanin[20], azurin[21], and plastocyanin[22] (Δβ data for azurin are shown in **Fig. S4**). The same

trends are noted, including low or near zero Δβ values for turn regions and gradual increases/decreases in Δβs for nearby residues in β-strands. Inspection of the azurin data (1JZF) shows that of its many defined turn regions, three (i.e., residues 10-13, 89-90, and 114-115) correspond to areas with Δβ values near zero (**Fig. 4B,** purple), as apparent from Δβ plots for n=9 and higher (**Fig. S4)**. Importantly, these three turn regions are thus the geometrical anchor points in azurin. Upon comparing the positions of the turns with Δβ values near zero in azurin and amicyanin (overlay shown in **Fig. S5**), we find two of the turns (the one after the first β-strand and the one before the last β-strand in the proteins' sequences) to be in the same position in the two structures; and both turns are nearby the Cu cofactor. The third turn with Δβ near zero in azurin corresponds to a turn with low Δβ in amicyanin. The third turn with Δβ near zero in amicyanin is positioned where azurin has an extended segment of two helices. Notably, the second of those two helices (which aligns sequence-wise with the mentioned zero Δβ turn in amicyanin) has residues with near zero Δβ (cyan, **Fig. S4)**.

To test the robustness and accuracy of our exact analytical approach to define protein structures in geometrical terms, several control calculations have been carried out. In **Fig. S6** we compare our amicyanin Δβ calculations using Cu as the origin of the coordinate system with those using the crystallographic origin as the external reference point (shown for three different n values). The results are quantitatively different, as expected, but the qualitative trends are the same. Importantly, the turn regions retain low or zero Δβ values when the crystallographic center is used as origin, and some β-strands show trailing Δβ behavior. In **Appendix** in the SI we show that if the origin of the coordinate system is moved to coincide with a residue on the backbone (á la Ramachandran), the angles calculated in two different (independent) ways are (sensibly) the same.

Before we close, we have more to say about unstructured regions in proteins. Hydrophobic indices and other thermodynamic scales often are used to define types of structures in different protein regions, such as the region that includes residues 47-75 in amicyanin. Although it has been designated as unstructured, our $\Delta\beta$ analysis shows that this region does not have a unimodal property with respect to geometry. Since attractive/repulsive forces govern the folded geometry, those forces, like $\Delta\beta$, are not uniform throughout this region and averaging is not appropriate. Nonetheless, inspection of this region from a geometrical perspective reveals which stretches (those with similar $\Delta\beta$ values) can be used for averaging.

Conclusions

We have developed an analytic approach that defines angular correlations between and among residues in folded protein structures. The method uses the spatial coordinates of each residue from a crystal structure exactly to calculate angular coordinates relative to a fixed point external to the protein backbone. The calculations involve only trigonometric functions, and they can be performed on a standard workstation in a few seconds. Our approach is illustrated in detail for the blue copper protein amicyanin with its copper ion cofactor acting as an external reference point. Importantly, we identified specific turn regions to act as symmetrical anchors in blue copper protein structures. Further studies of proteins with other folds (the blue copper proteins lack significant amount of α-helix) will reveal if symmetrical anchor points is a common feature of proteins or, if this is a specific property of blue copper proteins studied here. Since all that is needed to implement our method is a high-resolution structure, any one of the thousands of proteins that have been structurally characterized to date can be analyzed in this way. Our current analysis of α-helical heme proteins (cytochrome c, cytochrome b_{562} and

cytochrome c') implies that helices will have their own specific Δβ signatures (to be published elsewhere). As mentioned above, although a cofactor is a simple choice of external reference point, the crystallographic center (or any other convenient position) can be used as reference point for proteins without cofactors.

We propose that this type of geometrical analyses could produce datasets of common angular correlations for different secondary structure elements that in turn could be used as a refinement step in future design of proteins with specified functions.

Acknowledgements

We thank Istvan Horvath (Chalmers University) for making structural images. Roberto Garza-Lopez (Pomona College) provided helpful insights in early stages. Work performed at Caltech was supported by NIH (R01 DK019038 to HBG) and by the Wallenberg Foundation (funding a sabbatical for PWS at Caltech).

Supplemental Information. Figures S1-S6, and Appendix

References

1. W. Kauzmann, Some factors in the interpretation of protein denaturation. Adv. Protein Chem. 14, **1959**, 1-63.
2. J.J. Kozak, W.S. Knight, W. Kauzmann, Solute-solute interactions in aqueous solution. J. Chem. Phys., 48, **1968**, 675-690.
3. R.L. Baldwin, G.D. Rose, How the hydrophobic factor drives protein folding. Proc. Natl. Acad. Sci. (USA) 113, **2016**, 12462–12466.
4. C. Tanford, The hydrophobic effect: formation of micelles and biological membranes. New York, Wiley, 1973.
5. C. Tanford, Interfacial free energy and the hydrophobic effect, Proc. Natl. Acad. Sci. (USA) 76, **1979**, 4175-4176.
6. A.R. Henn (Editor). Festschrift in honor of Walter J. Kauzmann, Biophys. Chem. 105, **2003**, 153-774.
7. G.D. Rose, P.J. Fleming, J.R. Banavar, A. Maritan, A backboned-based theory of protein folding, Proc. Natl. Acad. Sci. (USA) 103, **2006**, 16623-16633.
8. M. Compiani, E. Capriotti, Computational and theoretical methods for protein folding, Biochemistry 52(48), **2013**, 8601-8624.
9. D. Chandler, Interfaces and the driving force of hydrophobic assembly, Nature, 437, **2005**, 640-647.
10. H.S. Chung, K. McHale, J.M. Louis, W.A. Eaton, Single molecule fluorescence experiments determine protein folding transition path times, Science, 335(6071), **2012**, 981-984.
11. V. Dagget, A. Fersht, The present view of the mechanism of protein folding, Nature Reviews Mol. Cell Biol. 4, **2003**, 487-502.
12. C. L. Brooks III, M. Gruebele, J.N. Onuchic, P.G. Wolynes, Chemical physics of protein folding, Proc. Natl. Acad. Sci. (USA) 95, **1998**, 11037-11038.
13. P.-S. S.E. Boyken, D. Baker, The coming of age of de novo protein design, Nature 537, **2016**, 320-327.
14. G.N. Ramachandran, C. Ramakrishnan, V. Sasisekharan, Stereochemistry of polypeptide chain configurations. J. Mol. Biol. 7, **1963**, 95-99.
15. G.N. Ramachandran, V. Sasisekharan, Conformation of polypeptides and poteins, Adv. Protein Chem. 23, **1968**, 283-438.
16. J.J. Kozak, H.B. Gray, R.A. Garza-Lopez, Relaxation of structural constraints during amicyanin unfolding, J. Inorg. Biochem. 179, **2018**, 135-145.
17. R. Durley, L. Chen, L.W. Lim, F.S. Mathews, V.L. Davidson, Crystal structure analysis of amicyanin and apoamicyanin from *Paracoccus denitrificans* at 2.0 Å and 1.8 Å resolution, Protein Sci., 2, **1993**, 739-752.
18. N. Sukumar, F.S. Mathews, P. Langan, V.L. Davidson, A joint X-ray and neutron study on amicyanin reveals the role of protein dynamics in electron transfer, Proc. Natl. Acad. Sci. (USA), 107(15), **2010**, 6817-6822.
19. L.M. Cunane, Z.W. Chen, R.C. Durley, F.S. Mathews, X-ray structure of the cupredoxin amicyanin, from *Paracoccus denitrificans*, refined at 1.31 Å resolution, Acta Crystallogr. D52, **1996**, 676-686.
20. R.L. Walter, S.E. Ealick, A.M. Friedman, R.C. Blake II, P. Proctor, M. Shoham, Multiple wavelength anomalous diffraction (MAD) crystal structure of rusticyanin; a highly oxidizing cupredoxin with extreme acid stability, J. Mol. Biol. 263, **1996,** 730-751.
21. B.R. Crane, A.J. Di Bilio, J.R. Winkler, H.B. Gray, Electron tunneling in single crystals of *Pseudomonas aeruginosa* azurins, J. Am. Chem. Soc. 123, **2001**, 11623-11631.
22. G.S. Kacholova, A.C. Shosheva, G.P. Bourenkov, A. Donchev, M. Dimitrov, H.D. Bartunik, Structural comparison of the poplar plastocyanin isoforms PC_a and PC_b sheds new light on the role of the copper site geometry in interactions with redox partners in oxygenic photosynthesis, J. Inorg. Biochem. 115, **2012**, 174-181.

TOC Graphics